INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

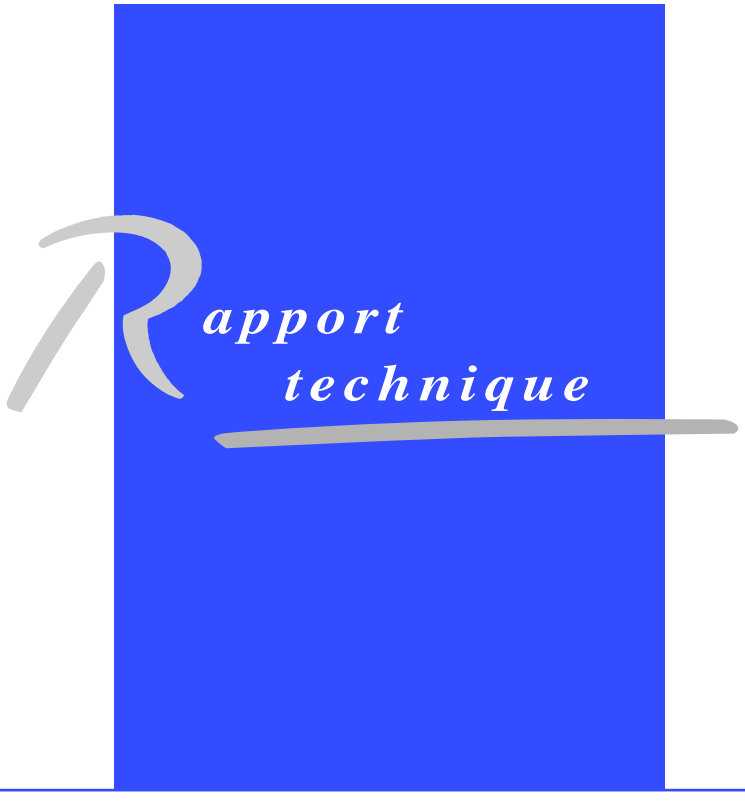
# *Towards Shape Constellation Inference through Higher-Order MRF Optimization in Nonlinear Embeddings*

Samuel Kadoury — Nikos Paragios

**N° 0376**

Décembre 2009

Thème BIO




ISSN 0249-0803 ISRN INRIA/RT--0376--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY - ÎLE-DE-FRANCE

# Towards Shape Constellation Inference through Higher-Order MRF Optimization in Nonlinear Embeddings

Samuel Kadoury*, Nikos Paragios*

Thème BIO — Systèmes biologiques
Équipe-Projet GALEN

Rapport technique n° 0376 — Décembre 2009 — 20 pages

**Abstract:** This paper introduces a novel approach for inferring articulated shape models from images. A low-dimensional manifold embedding is created from a training set of prior mesh models to establish the patterns of global shape variations. Local appearance is captured from neighborhoods in the manifold once the overall representation converges. Inference with respect to the manifold and shape parameters is performed using a Markov Random Field (MRF). Singleton and pairwise potentials measure the support from the data and shape coherence from neighboring models respectively, while higher-order cliques encode geometrical modes of variation in localized shape models. Optimization of model parameters is achieved using efficient linear programming and duality. The resulting model is geometrically intuitive, captures the statistical distribution of the underlying manifold and respects image support in the spatial domain. Experimental results on articulated bodies such as spinal column geometry estimation demonstrate the potentials of our approach.

**Key-words:** Image inference, statistical modeling, non-linear manifold, articulated model, Markov Random Field

* Laboratoire MAS, Ecole Centrale de Paris, Grande Voie des Vignes, 92295 Chatenay-Malabry, France

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

# Inférence de Modèles Articulés par une Optimisation MRF de Haut-Niveau dans des Domaines Non-Linéaires

**Résumé :** Ce papier introduit une nouvelle approche pour inférer des modèles articulés à partir d'images. Un domaine approximativement linéaire par monceaux est créé afin de réduire la complexité des données d'une base d'apprentissage de modèles *a priori*, et ce afin d'établir les patrons de variations. L'apparence locale est obtenue par une analyse de voisinages dans le domaine lorsque la représentation globale converge. L'inférence par rapport aux paramètres est effectuée par un Markov Random Field (MRF). Des valeurs potentiels unitaires et binômes mesurent le support avec les données et la cohérence de l'objet avec les modèles avoisinant respectivement, alors que des fonctions de haut niveau introduisent des contraintes géométriques pour les modes de variations locales. L'optimisation des paramètres est effectuée par une approche de programmation linéaire et par dualité. Le modèle résultant est intuitif géométriquement, capture la distribution statistique du sous-domaine et respecte les contraintes de l'image. Des résultats expérimentaux sur des formes articulés telles que la colonne vertébrale démontrent le potentiel de notre approche.

**Mots-clés :** Inférence d'images, modélisation statistique, domaines non-linéaires, modèle articulé, Markov Random Field

# Contents

**1 Introduction** **4**

**2 Manifold Embedding of Articulated Models** **6**
2.1 Representation of Articulated Models . . . . . 6
2.2 Nonlinear Manifold Embedding of ADMs . . . . . 6
2.2.1 Nearest neighbor selection . . . . . 6
2.2.2 Embedding algorithm . . . . . 8
2.2.3 Inverse mapping to the ambient space . . . . . 9
2.3 Local Shape Appearances in the Manifold . . . . . 9

**3 Inference through MRF Optimization** **11**
3.1 Rigid Alignment of the ADM . . . . . 11
3.2 Non-Rigid Adaptation of Local Shapes . . . . . 12
3.3 Energy Minimization . . . . . 14

**4 Experimental Validation** **15**

**5 Discussion** **17**

# 1 Introduction

Statistical models of shape variability have been successful in addressing fundamental vision tasks such as segmentation and registration in computer vision and medical imaging analysis. Such models help to understand the distribution in appearance of a group of shapes and offer efficient parametrization of the geometric variability in a given cluster. In medical imaging, one has the ability to compare information of a subject over a time period, between different subjects or between groups of pathological cases by using these types of representation [1]. These models have been used extensively for localized structures. On the other hand object constellations have been sparely investigated in computer vision (mainly on body [2] and hand-pose estimation [3]), which is mostly due to the challenges in interpreting the images.

Various statistical approaches for three-dimensional modeling of structures have been proposed based on shape analysis. Active Shape (ASM) [4] and Appearance Models (AAM) [5] have been successful in recovering object geometries obtained from dense collection of data points. Implicit representations is an alternative formulation [6] to address model-based segmentation while more recently numerous methods based on point distribution models and embedding on various geometric spaces (spherical [7]) have been proposed. However, model-based segmentation of single objects typically leads to fitting errors when no clear object boundary is present, similar structures in close vicinity, or pathologies. The result is thereby sensitive to model initialization and therefore limited to the capture range.

Simultaneous multi-object inference is often beneficial compared to the separate segmentation of individual objects. In [8], an extension of point distribution models (PDMs) was considered for modeling relations between shape constellations using conditional probabilities between 2D-contours. Furthermore, rigid transformations were considered as statistics between different parts to evaluate the variability from a given class of models [9] and to achieve multi-modal inference [10].

Last, but not least in the context of medical imaging, traditional body-pose estimation methods are prone fail. This is due to the fact that the aim of articulated models is to deduce the anatomical structure but most importantly to identify/segment the individual structures precisely, which is not the focus of vision methods. In vision where mostly 2D models are used, classification techniques are often considered to provide candidate positions for the different parts [11]. This information along with prior constraints are used to determine the most appropriate pose configuration.

The main limitation of the above mentioned methods is that incorporating image-driven information is not straightforward. Furthermore, the high dimensionality and complex non-linear underlying structure unfortunately makes the commonly used linear statistics [12] inapplicable for articulated structures. Given the important and often un-correlated variations in the orientation and translation of the global shape, such statistics can only be performed along a linear manifold that appropriately parameterizes the nonlinear space. Recent concepts on Riemannian manifolds in tensor spaces is particulary suited for this application [13].

A manifold learning algorithm of particular interest to this work is locally linear embedding [14]. It maps high-dimensional observation data that are presumed to lie on a nonlinear manifold, onto a single global coordinate system of lower dimensionality. Therefore, it helps to preserve neighborhood relationships of similar object geometries, thereby revealing the underlying structure of the data which can be used for statistical modeling. Inferring a model from the underlying manifold is a novel concept but far

from being trivial, and relies on a cost function which includes visual support and prior constraints.

In this paper, we introduce a deformable articulated body instantiation through a statistical modeling of inter-object transformations. We use nonlinear manifold embeddings which can handle both small and large deformations of the object constellations in a given dataset. Our principal contribution lies with regards to the representation of the model. Rather than directly optimizing a set of transformation parameters, we perform a statistical embedding on the transformation vectors which describe the articulations, thus significantly reducing the degrees of freedom. An explicit representation of the nonlinear manifold given by smooth functions enables the joint mapping between ambient and embedded space. Furthermore, because of the limited dimensionality of the search space, the inference with respect to the low-dimensional parameters is performed in an integrated and interconnected high-order MRF. This graph involves costs related both to the data, prior geometrical dependencies and global higher-order cliques. We consider a framework which applies recent advances in MRF labeling schemes generalizing energy function classes with degrees higher than two which can be decomposed into pair-wise potentials [15]. This demonstrates the capability to model complex interactions of random variables. Recent advances in the area of discrete optimization which explore duality theorem of linear programming [16] are explored to obtain the lowest potential of the objective function.

This paper is organized as follows. Section 2 presents the theoretical methodology for the manifold embedding algorithm of articulated mesh models while in Section 3 we propose the MRF-based inference framework minimizing $\mathcal{P}^n$ Potts models energy terms. In Section 4 we present our evaluation results applied to medical images of the spinal column and the last section concludes the paper.

# 2 Manifold Embedding of Articulated Models

The input to our method is a sample of articulated models which comprises a set of learning shapes. These shapes are a constellation of triangular meshes, each annotated with landmarks defined as characteristic points uniquely localized across a set of objects. We first build an articulated shape model from a training database by embedding the data into a low-dimensional sub-space which dimensionality corresponds to the domain of admissible variations. Local shape appearance is determined via analysis of variations within a sub-patch of the manifold.

## 2.1 Representation of Articulated Models

Our shape model $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_L\}$ consists of an interconnection of $L$ objects. For each local shape $\mathbf{s}_i$, we recover a triangular mesh with vertices $\{\mathbf{v}_j^i | j = 1, \ldots, V\}$, where the $j^{th}$ vertex corresponds to approximately the same surface location from one shape to another. Additionally, every $\mathbf{s}_i$ is annotated with landmarks on each personalized model to rigidly register each object to its upper neighbor. The resulting rigid transforms are stored for each inter-object link. These transforms can also be determined via an ICP-like algorithm to recover the extrinsic parameters. Hence, an articulated deformable model (ADM) is represented by a vector of local inter-object rigid transformations $A = [T_1, T_2, \ldots, T_L]$ as illustrated in [Fig. 1]. To perform global shape modeling of the shape $\mathbf{S}$, we convert $A$ into an absolute representation

$$\mathbf{A}_{\text{abs}} = [T_1, T_1 \circ T_2, \ldots, T_1 \circ T_2 \circ \ldots \circ T_L] \tag{1}$$

using recursive compositions. The transformations are expressed in the local coordinate system (LCS) of the lower object, which can be defined proper to the object's geometrical representation capturing main axes of deformation. Center of transformation is located at the midpoint of the triangular mesh. The rigid transformations described in this paper are the combination of a rotation matrix $R$, a translation vector $t$ and the scaling $s$. We formulate the rigid transformation $T = \{s, R, t\}$ of a triangular mesh model as $y = sRx + t$ where $x, y, t \in \Re^3$. Composition is given by $T_1 \circ T_2 = \{R_1 R_2, R_1 t_2 + t_1\}$, while inversion as $T^{-1} = \{R^T, -R^T t\}$.

## 2.2 Nonlinear Manifold Embedding of ADMs

Let us consider $N$ articulated shape models expressed by the absolute vector representation $\mathbf{A}_{\text{abs}}^i$, of dimensionality $D$. The aim is to create a low-dimensional manifold consisting of $N$ points $Y_i$, $Y_i \in \Re^d$, $i \in [1, N]$ where $d \ll D$. In such a framework, if an adequate number of data points is available, then the underlying manifold $\mathcal{M}$ is considered to be "well-sampled". Therefore, it can represent the underlying population structure. In the sub-cluster corresponding to a pathological population, each individual point of the training set and its neighbours would lie within a locally linear patch on the manifold.

### 2.2.1 Nearest neighbor selection

The main limitation of embedding algorithms is the assumption of Euclidean metrics in the ambient space to evaluate similarity between sample points. In our approach, we adopt the intrinsic nature of the Riemannian manifold geometry allowing us to discern between articulated shape deformations in a topological invariant framework. The

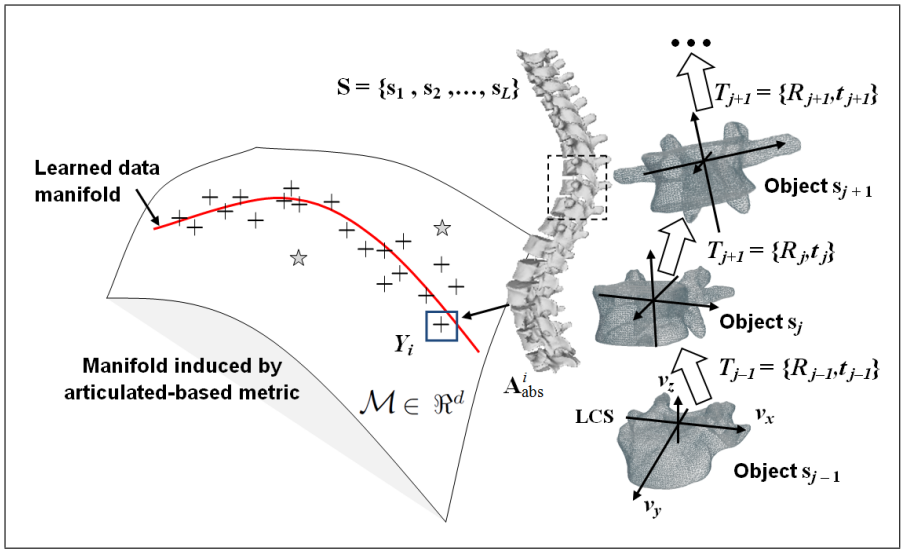



Figure 1: Schematic representation of inter-object transformations.

$K$ closest neighbours are selected for each point using a distortion metric particularly suited for geodesic metrics, defined as $d_{\mathbb{M}}(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}})$ which estimates the distance of articulated models $i, j$. As such, $\mathbf{A}^i_{\text{abs}}$ and $\mathbf{A}^j_{\text{abs}}$ are represented by the feature vectors described in (1). The distance measure can therefore be expressed as a sum of articulation deviations:

$$\begin{aligned} d_{\mathbb{M}}(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}}) &= \sum_{k=1}^{L} d_{\mathbb{M}}(T^i_k, T^j_k) \\ &= \sum_{k=1}^{L} d_{\mathbb{M}}\left( \begin{pmatrix} R^i_k & c^i_k \\ \dots & 1 \end{pmatrix}, \begin{pmatrix} R^j_k & c^j_k \\ \dots & 1 \end{pmatrix} \right) \end{aligned} \tag{2}$$

where the canonical representation encodes the intrinsic ($c$) and orientation ($R$) parameters. The difference between analogous articulations is computed within the geodesic framework:

$$d_{\mathbb{M}}(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}}) = \sum_{k=1}^{L} \|c^i_k - c^j_k\| + \sum_{k=1}^{L} d_Q(R^i_k, R^j_k). \tag{3}$$

The first term evaluates intrinsic distances in the $L_2$ norm. Using the geodesics, it is possible to define a diffeomorphism between rotation neighborhoods in $\mathcal{M}$ and a tangent plane $T_x\mathcal{M}$. The exponential map at $x \in \mathcal{M}$ maps vectors of the tangent plane $T_x\mathcal{M}$ to a point in the manifold which is reached by the geodesic $\gamma_{x,v}$ in a unit time. In other words, if $\gamma_{(x,v)}(1) = y$, then the inverse mapping is known as $\log_x(y) = v$. The quaternion distances are therefore computed with the Frobenius norm $d_Q(R^i_k, R^j_k) = \|\log((R^i_k)^{-1} R^j_k)\|_F$ based on the geodesic distance on the 3D manifold. This is feasible since rotations $R^i_k, R^j_k$ are nonsingular, invertible matrices. The $\log$ map can be thought as the unfolding operation connecting the tangent space at

$x$ and the manifold $\mathcal{M}$. One can now proceed to the manifold reconstruction using the local support in high-dimension data.

### 2.2.2 Embedding algorithm

The manifold reconstruction weights are estimated by assuming the local geometry of the patches can be described by linear coefficients that permit the reconstruction of every model point from its neighbours. In order to determine the value of the weights, the reconstruction errors are measured using the following objective function:

$$\min_{W} \varepsilon(W) = \sum_{i=1}^{N} \left\| \mathbf{A}^i_{\text{abs}} - \sum_{j=1}^{K} W_{ij} \mathbf{A}^j_{\text{abs}} \right\|^2 \tag{4}$$

$$\text{subject to } \begin{cases} W_{ij} = 0 \text{ if } \mathbf{A}^i_{\text{abs}} \text{ not neighbor } \mathbf{A}^j_{\text{abs}} \\ \sum_j W_{ij} = 1 \text{ for every } i, \end{cases} \tag{5}$$

Here, $\mathbf{A}^i_{\text{abs}}$ is the absolute vector describing an articulated model described above and $\varepsilon(W)$ sums the squared distances between all data points and their corresponding reconstructed points. The weights $W_{ij}$ represent the importance of the $j^{th}$ data point to the reconstruction of the $i^{th}$ element. It is easy to show that each weight can be calculated individually [14]. Each sample $\mathbf{A}^i_{\text{abs}}$ and weight $W_{ij}$ contributes to $\varepsilon$ such that:

$$\varepsilon^{(i)}(W) = \sum_{j=1}^{K} \sum_{m=1}^{K} W_{ij} W_{im} Q^i_{jm} \tag{6}$$

with $Q$ as a $K \times K$ matrix s.t. $Q^i_{jm} = (\mathbf{A}^i_{\text{abs}} - \mathbf{A}^j_{\text{abs}})^T (\mathbf{A}^i_{\text{abs}} - \mathbf{A}^m_{\text{abs}})$. With $\mathbf{R} = (\mathbf{Q} + \rho\mathbf{I})^{-1}$, given a suitably chosen regularization constant $\rho$ (see [17]), then $W_{ij}$ are solved by a least squares problem given the constraint in (5), $W_{ij} = (\sum_{m=1}^{K} R_{jm})/(\sum_{p,q=1}^{K} R_{pq})$. We propose to solve $Q$ based on distance matrix $D$ using articulated metric in (3) such that $Q = (D_{ij} + D_{im} - D_{jm})/2$.

The algorithm then maps each high-dimensional $\mathbf{A}^i_{\text{abs}}$ to a low-dimensional $Y_i$. These global internal coordinates are determined with a cost function minimizing the reconstruction error

$$\begin{aligned} \Phi(Y) &= \sum_{i=1}^{N} \left\| Y_i - \sum_{j=1}^{K} W_{ij} Y_j \right\|^2 \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij} Y_i^T Y_j \end{aligned} \tag{7}$$

with $\mathbf{M}$ as a sparse and symmetric $N \times N$ matrix enclosing the reconstruction weights $W_{ij}$ such that $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$, and $\mathbf{Y}$ spanning the $Y_i$'s. By constraining $\mathbf{Y}$ to identity, the problem becomes a straightforward one with $\min \text{tr}(\mathbf{YMY}^T)$ given the constraint $\frac{1}{N}\mathbf{YY}^T = \mathbf{I}$. Using Lagrange multipliers and setting the derivative to zero gives $(\mathbf{M} - \Lambda)\mathbf{Y}^T = 0$, where $\Lambda$ is the diagonal Lagrange multiplier matrix. The optimal embedding, up to a global rotation, is obtained from the bottom $d + 1$ eigenvectors of the and helps to minimize the cost function $\Phi(Y)$ as a simple

eigenvalue problem. The $d$ eigenvectors form the $d$ embedding coordinates. The coordinates $Y_i$ can be translated by a constant displacement without affecting the overall cost $\Phi(Y)$. The eigenvector corresponding to the smallest eigenvalue corresponds to the mean value of the embedded data $Y^0 = \{y_1, \ldots, y_d\}$, s.t. $y_i = 0, \forall i$. This can be discarded with $\sum Y_i = 0$ to obtain an embedding centered at the origin. Hence, a new ADM denoted as $Y_{\text{new}}$ can be inferred in the embedded $d$-space as a low-dimensional data point by finding its optimal manifold coordinates $y_i$.

### 2.2.3 Inverse mapping to the ambient space

To obtain the articulation vector for a new embedded point in the ambient space (image domain), one has to determine the representation in high-dimensional space based on its intrinsic coordinates. We first assume an explicit mapping $f : \mathcal{M} \rightarrow \Re^D$ from manifold $\mathcal{M}$ to the ambient space $\Re^D$. The inverse mapping of $Y_i$ is then performed by estimating the relationship between $\Re^D$ and $\mathcal{M}$ as a joint distribution, such there exists a smooth functional which belongs to a local neighborhood. Theoretically the manifold should follow the conditional expectation:

$$f(Y_i) \equiv E(\mathbf{A}^i_{\text{abs}} | \mathcal{M}(A_i) = Y_i) = \int A_i \frac{f(Y_i, A_i)}{f_{\mathcal{M}(A_i)}(Y_i)} dA \tag{8}$$

which captures the overall trend of the data. Here, both $f_{\mathcal{M}(A_i)}(Y_i)$ (marginal density of $\mathcal{M}(A_i)$) and $f(Y_i, A_i)$ (joint density) are unknown. Using *Nadaraya-Watson* kernel regression theorem [18], densities are replaced by kernel functions based in a conditional expectation setting [19]. We replace these variable with $f_{\mathcal{M}(A_i)}(Y_i) = \frac{1}{K} \sum_{j \in \mathcal{N}(i)} G_h(Y_i, Y_j)$ and $f(Y_i, A_i) = \frac{1}{K} \sum_{j \in \mathcal{N}(i)} G_h(Y_i, Y_j) G_g(A_i, A_j)$. The Gaussian regression kernels $G$ require the neighbors $\mathbf{A}^j_{\text{abs}}$ of $j \in \mathcal{N}(i)$ to determine the bandwidths $h, g$ so it includes all $K$ data points. Plugging these estimates in (8), this gives:

$$f_{\text{NW}}(Y_i) = \int A_i \frac{\frac{1}{K} \sum_{j \in \mathcal{N}(i)} G_h(Y_i, Y_j) G_g(A_i, A_j)}{\frac{1}{K} \sum_{j \in \mathcal{N}(i)} G_h(Y_i, Y_j)} dA. \tag{9}$$

By assuming $G$ is symmetric about the origin and generalize the expectation such that the observations $Y$ are defined in terms of a metric $d_{\mathbb{M}}$ in manifold space $\mathcal{M}$, we obtain:

$$f_{\text{NW}}(Y_i) = \underset{\mathbf{A}^i_{\text{abs}}}{\text{argmin}} \frac{\sum_{j \in \mathcal{N}(i)} G(Y_i, Y_j) d_{\mathbb{M}}(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}})}{\sum_{j \in \mathcal{N}(i)} G(Y_i, Y_j)} \tag{10}$$

which integrates the distance metric $d_{\mathbb{M}}(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}})$ defined in (3) and updates $f_{\text{NW}}(Y_i)$ using the closest neighbors of point $Y_i$ in the manifold space. This constrains the regression to be valid for similar data points in its vicinity since locality around $Y_i$ preserves locality in $\mathbf{A}^i_{\text{abs}}$.

## 2.3 Local Shape Appearances in the Manifold

The key idea of capturing local shape appearance lies on the assumption that models, represented in a given neighborhood $\mathcal{M}_k$ of the general manifold $\mathcal{M}$ s.t. $\mathcal{M}_k \in \mathcal{M} : \Re^d$, will also manifest similar local geometries. We assume here that local shape appearances follow a linear distribution within the low-dimensional manifold. Hence, given a data point $Y_j$ and its $K$ neighbors, the local shape model $\mathbf{s}_i$, representing the

$i^{th}$ element of the ADM, is obtained by building a particular class of shapes given the set of examples $\{\mathbf{s}_i^1, ..., \mathbf{s}_i^K\}$. This set belongs to the sub-patch $\mathcal{M}_k$, such that each shape $\mathbf{s}_i^j \in \mathbf{S}^j$. We approximate the distribution of the shape using a parameterized linear model by computing the deformation vectors formed for the $K-1$ shape samples. We compute the $n$ eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and corresponding eigenvectors, $v_1, v_2, \ldots, v_n$ of the $i^{th}$ covariance matrix:

$$C_i = \frac{1}{K-1} \sum_{j=1}^{K} (\mathbf{s}_i^j - \bar{\mathbf{s}_i})(\mathbf{s}_i^j - \bar{\mathbf{s}_i})^{\mathrm{T}} \tag{11}$$

where $\bar{\mathbf{s}_i}$ is the mean shape of the $K$ neighboring local objects $\bar{\mathbf{s}_i} = \frac{1}{K} \sum_{j=1}^{K} \mathbf{s}_i^j$. By arranging eigenvalues in increasing order, a new instance $\mathbf{s}_i^{\text{new}}$ of the shape model can be defined. The new shape is approximated by a linear combination of the deformation vector as follows:

$$\begin{aligned} \mathbf{s}_i^{\text{new}} &\approx \bar{\mathbf{s}_i} + \sum_{i=1}^{n} \omega_i v_i = \bar{\mathbf{s}_i} + [v_1 \ldots v_n][\omega_1 \ldots \omega_n] \\ &= \bar{\mathbf{s}_i} + \mathbf{V}\mathbf{w} \end{aligned} \tag{12}$$

where $\mathbf{V}$ is a $K \times n$ matrix with eigenvectors in each column, and $\mathbf{w} = [\omega_1 \ldots \omega_n]$ the weight vector of size $n$. In (12), $\mathbf{w}$ provides a compact parametrization of the transformation we seek for each object composing the model shape $\mathbf{S}$ based on the principal deviations. This final step warps individual instances from the training set to infer new local shape models.

# 3 Inference through MRF Optimization

Once an appropriate modeling of shape variations is determined, a successful inference between the image and manifold must be accomplished. We describe here how a new model is deformed, the similarity criterions as well as the adopted optimization procedure which infers the shape model to the sparse data. We search the optimal embedded manifold point $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_d)$ of the global ADM. Such a strategy offers an ideal compromise between the prior constraints, as well as the individual shape variation described by the main weight vectors $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ in a localized sub-patch. Formally, the inference of the model $\mathbf{S}$ to the image $\mathcal{I}$ is given by:

$$(\{\mathbf{y}_1, \ldots, \mathbf{y}_d\}; \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}) = \underset{\vec{\delta}_i, \omega_i}{\operatorname{argmin}} \; E(\mathbf{S}^0, \mathcal{I}, \vec{\Delta}, \Omega). \tag{13}$$

The energy $E$ of inferring the model $\mathbf{S}$ in the image $\mathcal{I}$ is a function of the displacement vectors $\vec{\Delta} = (\vec{\delta}_1, \ldots, \vec{\delta}_d)$ in the non-linear manifold space for global shape representation. This involves a data-related term $V(\mathbf{Y}^0 + \vec{\Delta}, \mathcal{I})$ expressing the image cost and a global prior term $V(\mathbf{N}, \vec{\Delta})$ measuring deformation between low-dimensional vectors with shape models. Furthermore we introduce a higher-order term $V(\mathbf{H}, \vec{\Delta}, \Omega)$ which is expressed by the reconstruction weights $\Omega = (\omega_1, \ldots, \omega_n)$ for local shape modeling in a linear space. The energy function $E$ can therefore be decoupled to a global and local optimization scheme controlled by the weighting parameters $\alpha$-$\beta$:

$$\begin{aligned} E\Big(\mathbf{S}^0, \mathcal{I}, \vec{\Delta}, \Omega\Big) = V\Big(\mathbf{Y}^0 + \vec{\Delta}, \mathcal{I}\Big) + \alpha \; V\Big(\mathbf{N}, \vec{\Delta}\Big) \\ + \beta \; V\Big(\mathbf{H}, \vec{\Delta}, \Omega\Big). \end{aligned} \tag{14}$$

We explain in this section how we define the global and local energy terms.

## 3.1 Rigid Alignment of the ADM

The global alignment of the model with the target (i.e., image data) primarily drives the deformation of the ADM in the first phase of convergence controlled by the $\alpha$-term such that it is minimized when the neighbors $K$ in a manifold sub-patch stabilizes. The purpose here is to estimate the set of articulations describing the global model shape by determining its optimal representation $\mathbf{Y}^0$ in the embedded sub-space. This is performed by reformulating the global representation using the inverse mapping in (10) so $f_{\text{NW}}(Y_i + \vec{\Delta}) = f_{\text{NW}}(\{y_1 + \vec{\delta}_1, \ldots, y_d + \vec{\delta}_d\}) = \mathbf{A}^i_{\text{abs}} + \vec{D}$ with $\vec{D}$ as deformations in ambient space $\Re^D$. This allows to represent the model in image space based on its manifold space coordinates of $\mathcal{M}$. The global cost is expressed as:

$$\begin{aligned} V\Big(\mathbf{Y}^0 + \vec{\Delta}, \mathcal{I}\Big) &= V\Big(f_{\text{NW}}(\{y_1 + \vec{\delta}_1, \ldots, y_d + \delta_d\}), \mathcal{I}\Big) \\ &= V\Big(\mathbf{A}^0_{\text{abs}} + \vec{D}, \mathcal{I}\Big). \end{aligned} \tag{15}$$

Since the transformations $T_i$ are implicitly modeled in the absolute representation $\mathbf{A}^0_{\text{abs}}$, we can formally consider the singleton image-related term as a summation of costs

associated with each $L$ objects of the ADM:

$$V\Big(\mathbf{A}^0_{\text{abs}} + \vec{D}, \mathcal{I}\Big) = \sum_{i=1}^{L} V_i\Big(\mathbf{s}_i * (T^0_i + \vec{d}_i), \mathcal{I}\Big) \tag{16}$$
$$\text{where } V_i(\mathbf{s}, \mathcal{I}) = \sum_{\mathbf{v}_i \in \mathbf{s}} \mathbf{n}^T_i(\mathbf{v}_i) \nabla \mathcal{I}(\mathbf{v}_i),$$

is a modular term seeking to minimize the distance between the mesh vertices of the inferred ADM and image data $\mathcal{I}$ by a rigid transformation of the vectors. This globally modifies the mesh vertices in order to align them with the $\mathcal{I}$. In (16), $\mathbf{v}_i$ is the set of triangle centers, $\mathbf{n}_i$ is the triangle normal pointing outwards, and $\nabla\mathcal{I}(\mathbf{v}_i)$ is the image gradient at location $\mathbf{v}_i$. This modular term effectively measures the strength of the edges over the triangles corresponding to the inferred model. The prior constraint for the rigid alignment is defined in the manifold space by modeling pairwise potentials between low-dimensional features $y_i$ represented by the second term in (14):

$$\alpha\, V\Big(\mathbf{N}, \vec{\Delta}\Big) = \alpha_{ij} \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y^0_i + \vec{\delta}_i, y^0_j + \vec{\delta}_j). \tag{17}$$

This potential measures the distance between pairs $y^0_i$ and $y^0_j$ from the current data point coordinates to a prior distribution built with the $K$ points $y^{1,..,K}_i, y^{1,..,K}_j$ in the local neighborhood $\mathcal{M}_k$. This term represents the smoothness term of the global cost function to ensure that the deformation $\vec{\delta}_i$ applied to point coordinates is regular in the non-linear vicinity of variations. This is measured by the geodesic distance in the Riemannian manifold

$$V_{ij} = P\Big(\|(y^0_i + \vec{\delta}_i) - (y^0_j + \vec{\delta}_j)\|_F, \mathcal{N}(i)\Big). \tag{18}$$

The Frobenius distance is measured to a density function determined from the the $K$ samples in the neighborhood such that $P(x, \mathcal{N}) = \frac{1}{\sigma_{\mathcal{N}}\sqrt{2\pi}}(1 - e^{(-x-\mu_{\mathcal{N}})^2/2\sigma^2_{\mathcal{N}}})$ assigns a cost based on the probability that $x$ belongs to a Gaussian distribution, with $\mu$ and $\sigma$ calculated from the $i$ and $j$ vectors of $\mathcal{N}(i) \in \mathcal{M}_k$.

## 3.2 Non-Rigid Adaptation of Local Shapes

Local shape geometry for each of the ADM's components is obtained by varying the weight parameters of the principal variations. We parameterize these potentials with clique variables $\mathbf{w}_c$ taking on corresponding costs $\theta_q$ if the cliques are given to the weight vectors $\omega_c$. Hence the third term of (14) is described as a high-order functional:

$$\beta\, V\Big(\mathbf{H}, \vec{\Delta}, \Omega\Big) = \beta_c \sum_{c \in \mathcal{C}} V_c(\mathbf{w}^0_c + \omega_c) \tag{19}$$

where independent clique variables $c$ are treated as a graph minimization problem. Our prior term is represented by higher-order potentials of degree $n$, based on the eigenvalues in 2.3 of the $L$ local objects from our model $\mathbf{S}$. The potential functions are defined as:

$$V_c(\mathbf{w}^0_c) = \min\{\min_{q \in \{1,2,..,t\}} \theta_q + \Delta_q(\mathbf{w}^0_c), \theta_{\max}\} \tag{20}$$

with $\Delta_q(\mathbf{w}_c^0) = \sum_{i \in c; l \in \mathcal{L}} \psi_{il}^q \kappa(\omega_i = l)$ is a deviation function to evaluate the assignment of $\mathbf{w}_c^0$. The Kronecker delta function $\kappa$ generates binary variables, while the weights are assigned such that $\psi_{il}^q = \{0, \theta_{\text{max}}\}$ depending on whether the clique variable $\mathbf{w}_c$ is given the appropriate label.

Our work is inspired from a mesh reconfiguration where the costs $\theta_q$ depend on the positions of the morphed mesh vertices $\mathbf{v}_i$. It introduces a tradeoff between adherence to the image and coherence with regards to the prior principal modes of shape variations in the neighborhood $\mathcal{M}_k$. The strategy finds the point in the target image with an optimal boundary criterion. Hence for each mesh triangle, a search along the normal $\mathbf{n}_i$ from $\mathbf{v}_i$ with optimal compromise between boundary detection and distance $j\gamma$ is performed. It measures the cost of driving the triangle centers towards detected features points in the target image [20]:

$$\theta_q = \sum_{i=1}^{V} w_q \Big( \frac{\nabla \mathcal{I}(\mathbf{v}_i)}{\|\nabla \mathcal{I}(\mathbf{v}_i)\|} \gamma \mathbf{n}_i \underset{j=-m,\ldots,-m}{\operatorname{argmin}} (j^2\gamma^2 - \mathbf{n}_i^T \nabla \mathcal{I}(\mathbf{v}_i + j\gamma \mathbf{n}_i)) \Big)^2. \tag{21}$$

In (21), $V$ is the number of triangles, $m$ is the sample space to find the closest detected feature in $\nabla \mathcal{I}$ and $\gamma$ specifies the distance between two points on the normal. The weights $w_q$ controls the tradeoff of how far the directed gradient distance lies from the mean eigenvalue shape.

One can integrate the global data and prior terms along with local shape terms parameterized as the higher-order cliques, by combining (15), (17) and (19):

$$\begin{aligned} E\Big(\mathbf{S}^0, \mathcal{I}, \vec{\Delta}, \Omega\Big) = V\Big(f_{\text{NW}}(\{y_1 + \vec{\delta}_1, \ldots, y_d + \vec{\delta}_d\}), \mathcal{I}\Big) \\ + \alpha_{ij} \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^0 + \vec{\delta}_i, y_j^0 + \vec{\delta}_j) \\ + \beta_c \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^0 + \omega_c). \end{aligned} \tag{22}$$

### 3.3 Energy Minimization

The optimization strategy of the resulting MRF (22) in the continuous domain is not a straightforward problem. The convexity of the solution domain is not guaranteed, while gradient-descent optimization approaches are prone to non-linearity and local minimums. We therefore considered results obtained from discrete optimization approaches [16]. By assuming the initial zero-mean shape model in the given manifold space $\mathcal{M}$ falls in the image domain $\mathcal{I}$ and that the desired accuracy is specified so to define the quantization step, we can approximate the deformation of the shape model towards the optimal solution.

We seek to assign the optimal labels $\mathcal{L}^{\vec{\Delta}} = \{l_1, \ldots, l_d\}$ and $\mathcal{L}^{\Omega} = \{l_1, \ldots, l_n\}$ which are associated to the quantized space $\vec{\Delta}$ of displacements and local weight parameters $\Omega$. If we consider that displacing the coordinates of the sub-domain point $y_i^0$ by $\vec{\delta}^{l_i}$ is equivalent to assigning label $l_i$ to $y_i^0$, we can formulate the energy as:

$$(\{\mathbf{l}_1^{\vec{\Delta}}, \ldots, \mathbf{l}_d^{\vec{\Delta}}\}; \{\mathbf{l}_1^{\Omega}, \ldots, \mathbf{l}_n^{\Omega}\}) = \underset{l_i^{\Delta}, l_i^{\Omega} \in \mathcal{L}^{\Delta}, \mathcal{L}^{\Omega}}{\operatorname{argmin}} E(\mathbf{S}^0, \mathcal{I}, \mathcal{L}^{\vec{\Delta}}, \mathcal{L}^{\Omega}). \tag{23}$$

The quantization domain of the label set is crucial to achieve both inference accuracy and computational efficiency. To this end, we adopt a coarse-to-fine approach which continuously increases the number of displacements while decreasing the search space. This speeds up the convergence and helps to reduce the interaction between graph nodes. Furthermore to account for previously searched labels, an incremental approach is used where in each iteration $t$ we look for the set of labels that improves the current solution s.t. $y_i^t = y_i^0 + \sum_t \vec{\delta}^{l_i t}$, which is a temporal minimization problem. Then (22) can be re-written as a labeling problem:

$$\begin{aligned} E^t(\mathcal{L}^{\vec{\Delta}}, \mathcal{L}^{\Omega}) = V\Big(f_{\text{NW}}(\{y_1^{t-1}, l_1^{\vec{\Delta}}, \ldots, y_d^{t-1}, l_d^{\vec{\Delta}}\}), \mathcal{I})\Big) \\ + \alpha_{ij} \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^{t-1}, y_j^{t-1}, l_i^{\vec{\Delta}}, l_j^{\vec{\Delta}}) \\ + \beta_c \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^{t-1}, l_c^{\Omega}). \end{aligned} \tag{24}$$

We solve the minimization of the higher-order cliques in (24) by transforming them into quadratic functions [15] using a $(t+1)$-state switching variable which finds the deviation function which assigns the lowest cost to the labeling:

$$\min V_c(\mathbf{w}_c^0) = \min_{\mathbf{w}_c^0, z \in \{1,2,\ldots,t+1\}} f(z) + \sum_{i \in c} g(z, w_i) \tag{25}$$

where $f(z) = \{\theta_q, \theta_{max}\}$ is a cost assigning function depending on the state variable $z$ and $g(z, w_i) = \psi_{il}^q$ when $z = q$ and $w_i = l \in \mathcal{L}^{\Omega}$, while $g(z, w_i) = 0$ when $z = t+1$.

We apply a Primal-Dual algorithm called FastPD [16] which can efficiently solve the inference problem in a discrete domain by formulating the duality theory in linear programming. The advantage of such an approach lies in its generality, efficient computational speed, and guarantees the global optimum without the condition of linearity.

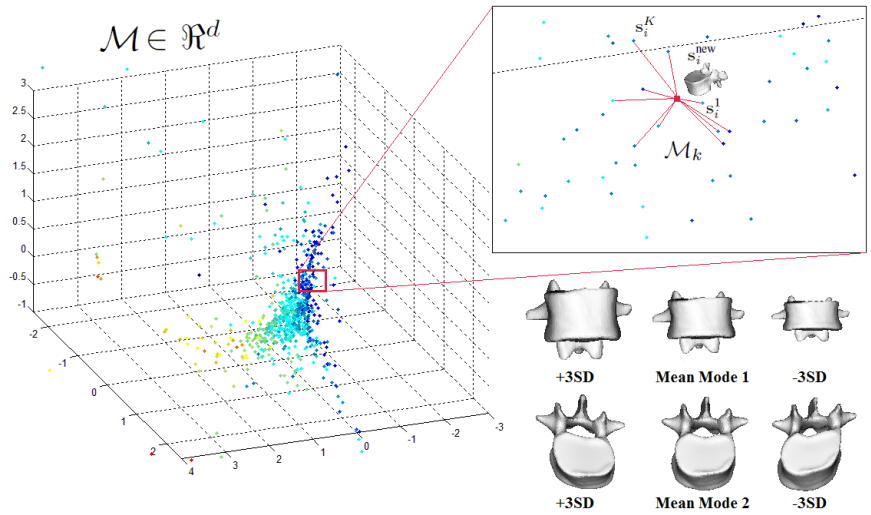
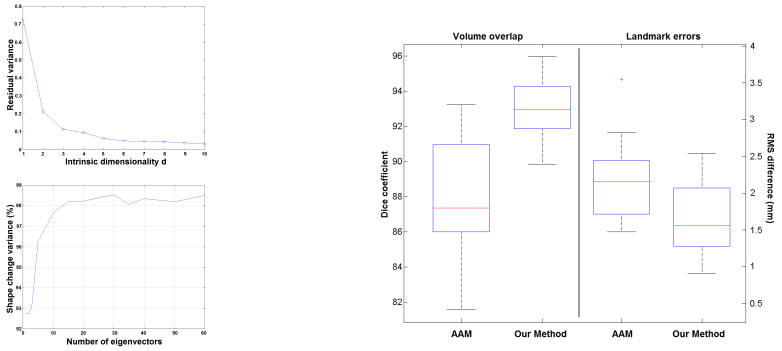



Figure 2: Manifold embedding result of the spine dataset with illustration of a local patch $\mathcal{M}_k$ for individual shape warping of two first modes $n = 2$ with variation of $\pm$3SD.

# 4 Experimental Validation

In order to validate the performance of the method, we considered modeling pathological spinal columns represented as articulated shapes for computerized tomography (CT) inference. We used a database of 711 spine models reconstructed in 3D from X-ray images, exhibiting different types of deformations relative to global and local shape geometries. For each spine, 6 landmarks on each of the 17 vertebrae composing the spinal column where used to define the LCS in order to extrapolate the inter-object transformation as described in Section 2.1. Models were composed between 3831 and 6942 vertices depending on the vertebra level. Dimensionality reduction provides the embedding manifold shown in [Fig. 2] with local variation mode from a selected patch. Optimal neighborhood size was found at $K = 10$, while intrinsic dimensions was $d = 7$ and $n = 5$ as shown in [Fig. 3(a)] dictating the number of nodes in our global graph model. We tested the algorithm on a subset of 20 unseen cases from the database. Successful examples of the obtained results from sparse sample points and CT inferred data are shown in [Fig. 4]. Additional results are presented in [Fig. 5]. Towards checking the robustness of the method, we evaluated the performance based on the density of the input sample points. [Fig. 6(a)] demonstrates the model's accuracy with respect to the number of available points, visibly affecting the global shape when density is reduced to 20%. To further validate the robustness, we added gaussian distributed noise to the target data. While distortions begin to appear when $\sigma = 3$, the inferences obtained in [Fig. 6(b)] are promising by simply varying the prior weight complements $\alpha$ and $\beta$. The method also performs well when local objects are completed occluded or missing [Fig. 6(c)]. We quantitatively compared our method to an AAM modeling based on global PCA [Fig. 3(b)]. Dice scores and root-mean-square (RMS) landmark distances computed from the test cases show improvement of the proposed MRF approach via a non-linear shape analysis. To demonstrate the potential of




(a) (b)

Figure 3: (a) Dimensionality of the global manifold based on residual variance (top) and of local shape eigenvalues using the % of covered shape variance (bottom);(b) Boxplots of dice overlap coefficients and landmark errors comparing our method to AAM.

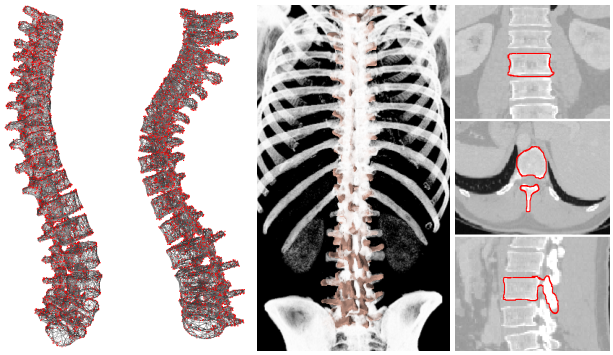
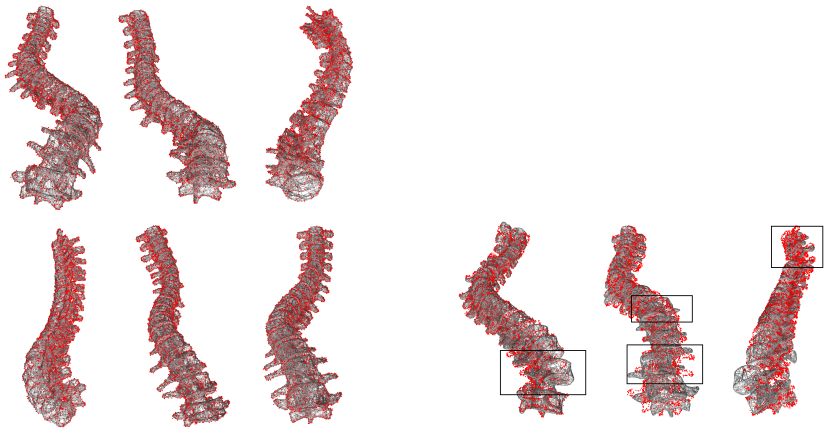


Figure 4: Successful spine inference results of triangulated mesh ADMs to sparse (red points) and CT image with orthogonal views.

our method for other modeling applications, we produced results on deformable arm poses (GRAIL, Univ. Washington) via an articulated representation shown in [Fig. 7]. These validations prove how our method elegantly encodes prior knowledge with image constraints in an MRF framework, and efficiently minimizes the energy term to converge towards an optimal solution. One drawback remains the computational time due to the inverse regression mapping and higher-order clique potential minimization. Experiments were conducted in C++ on a 2.8GHz Intel P4 processor and 4GB memory.



(a) (b)

Figure 5: (a)-(b) Examples of successful and erroneous inference results in various configurations.

# 5 Discussion

We have proposed a novel method for inferring articulated shape models. Our main contribution consists in modeling complex, non-linear patterns of prior deformations in a Riemannian manifold embedding. New point-based models are created from statistical knowledge in terms of global and local variations. To this end, we introduced a conditional regression kernel to perform the inverse mapping to the ambient space from neighbors selected with an articulated distance metric based on intrinsic and orientation properties. One observation of the non-linear embedding is that it avoids creating shape distortions and eliminates the need to solve large-dynamic programming problems, thus saving computational time and memory space. The proposed framework based on higher-order MRFs is general, and can be extended for other applications in vision and medical imaging to accommodate for pose modeling. Real-time inference of articulated models, motion tracking and medical guidance based on optimal higher-order clique decomposition would be beneficial towards this end.

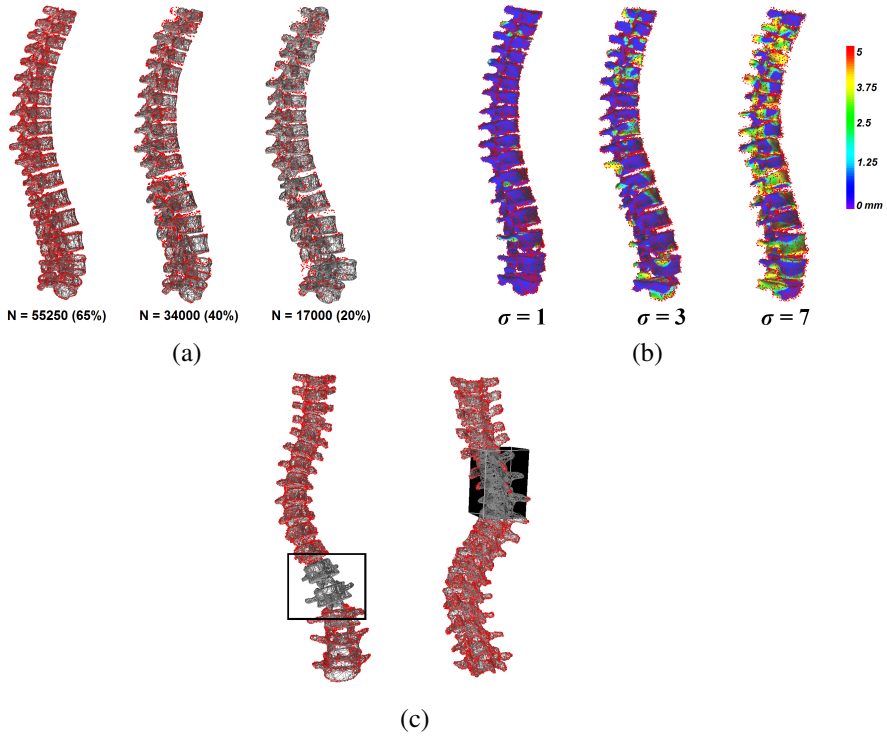



Figure 6: Triangulated mesh models represent inferred ADM; red points illustrate the target image data. (a) Decreasing number of available points to infer the ADM affecting global shape coherence. (b) Local shape distortions with significant noise level increase $\sigma$ added to target points (error-coded models for ground-truth distances). (c) Robustness of method towards missing target components and occluded parts.

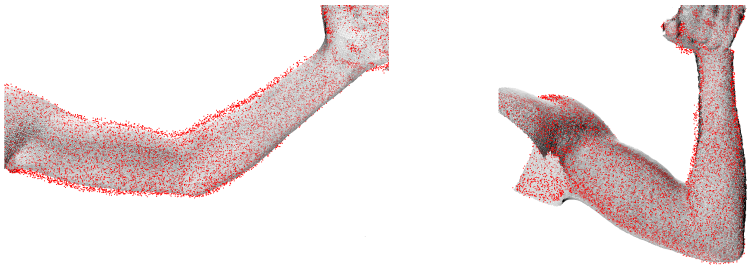


Figure 7: Human body part deformation results for the arm.

# References

[1] Rueckert, D., Frangi, A. and Schnabel, J.: Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. IEEE T. Med. Imag. **22** (2003) 1014-1025.

[2] Andriluka, M., Roth, S. and Schiele, B.: Pictorial structures revisited: People detection and articulated pose estimation. In: Proc. CVPR. (2009) 1014-1021.

[3] de La Gorce, M., Paragios, N. and Fleet, D.J.: Model-based hand tracking with texture, shading and self-occlusions. In: Proc. CVPR. (2008) 1-8.

[4] Cootes, T., Taylor, C., Cooper, D. and Graham, J.: Active shape models - their training and application. CVIU. **61** (1995) 38-59.

[5] Cootes, T., Edwards, G. and Taylor, C.: Active appearance models. PAMI. **23** (2001) 681-685.

[6] Rousson, M. and Paragios, N.: Prior knowledge, level set representations and visual grouping. IJCV. **76** (2008) 231-243.

[7] Nain, D., Haker, S., Bobick, A.F. and Tannenbaum, A.: Multiscale 3-D shape representation and segmentation using spherical wavelets. IEEE T. Med. Imag. **26** (2007) 598-618.

[8] de Bruijne, M., Lund, M., Tanko, L., Pettersen, P. and Nielsen, M.: Quantitative vertebral morphometry using neighbor-conditional shape models. In: Proc. MICCAI. (2006) 1-8.

[9] Boisvert, J., Cheriet, F., Pennec, X., Labelle, H. and Ayache, N.: Geometric variability of the scoliotic spine using statistics on articulated shape models. IEEE T. Med. Imag. **27** (2008) 557-68.

[10] Kadoury, S. and Paragios, N.: Surface/volume-based articulated 3D spine inference through Markov Random Fields. In: Proc. MICCAI. (2009) 92-99.

[11] Felzenszwalb, P. and Huttenlocher, D.: Pictorial structures for object recognition. IJCV. **61** (2005) 5579.

[12] Cootes, T., Breston, C., Edwards, G. and Taylor, C.J.: A unified framework for atlas matching using active appearance models. In: Proc. IPMI. (1999) 322-333.

[13] Khurd, P., Baloch, S., Gur, R., Davatzikos, C. and Verma, R.: Manifold Learning Techniques in Image Analysis of High-dimensional Diffusion Tensor Magnetic Resonance Images. In: Proc. CVPR. (2007) 1-7.

[14] Roweis, S. and Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science. **290** (2000) 2323-2326.

[15] Rother, C., Kohli, P., Feng, W. and Jia, J.: Minimizing sparse higher order energy functions of discrete variables. In: Proc. CVPR. (2009) 1382-89.

[16] Komodakis, N., Tziritas, G. and Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. CVIU. **112** (2008) 14-29.

[17] de Ridder, D. and Duin, R.P.W.: Locally linear embedding for classification. Tech. Report PH-2002-01. Delft University of Technology, Delft, Netherlands. (2002).

[18] Nadaraya, E.A.: On estimating regression. Theory of Probability and its Applications. **10** (1964) 186-190.

[19] Davis, B., Fletcher, P., Bullitt, E. and Joshi, S.: Population shape regression from random design data. In: Proc. ICCV. (2007) 1-7.

[20] Kaus, M., Pekar, V., Lorenz, C., Truyen, R., Lobregt, S. and Weese, J.: Automated 3-D PDM construction from segmented images using deformable models. IEEE T. Med. Imag. **22** (2003) 1005-1013.

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-0803